International Journal of Instruction *January 2027 ● Vol.20, No.1*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 317-332

Article submission code: 20260325133438
Received: 25/03/2026 Revision: 25/06/2026
Accepted: 06/07/2026 OnlineFirst: 01/10/2026

AI-Generated Feedback and Design Performance: The Role of Self-Efficacy in a Quasi-Experimental Study

Wenhao Dai
Rajamangala University of Technology Krungthep, Thailand & Chongqing Polytechnic University of Electronic Technology, China, *daiwenhao@cqcet.edu.cn*

Alan Robert White
Rajamangala University of Technology Krungthep, Thailand, *alan.w@mail.rmutk.ac.th*

Xingting Zhou
Rajamangala University of Technology Krungthep, Thailand, *979189791@qq.com*

Whether an AI-assisted feedback system providing structured, criterion-referenced commentary throughout the project cycle is associated with stronger outcomes in design education has remained underexamined. This quasi-experimental study compared two intact third-year advertising design classes at a Chinese university (N = 77), with one class receiving teacher-led critique and the other using an AI-supported iterative feedback system. The AIFF condition combined AI-mediated commentary with greater access to on-demand feedback and more opportunities for iterative revision. Across a 16-week semester, outcomes were measured using an 8-criterion design quality rubric (ICC = 0.89), the Consensual Assessment Technique for creative performance (ICC = 0.86), and a 12-item design self-efficacy scale (alpha = 0.84). ANCOVA controlling for pretest performance showed higher adjusted posttest scores for the AIFF group on design quality, creative performance, and self-efficacy, with descriptive between-group effect sizes of d = 0.77, d = 0.72, and d = 0.60, respectively; all group effects were significant at p <= 0.008. Within the AIFF group, iteration frequency was positively associated with design quality gains. These findings are consistent with the interpretation that a feedback condition combining AI mediation, greater feedback availability, and iterative revision opportunities may support improved learning outcomes in creative education. However, because of the quasi-experimental design, the specific effects of AI, feedback frequency, novelty, and student motivation cannot be disentangled. Implications for integrating AI-supported feedback into design studio pedagogy are discussed.

Keywords: AI-supported feedback, teacher-led critique, advertising design education, creative performance, design self-efficacy, iterative revision, quasi-experimental

Citation: Dai, W., White, A. R., & Zhou, X. (2027). AI-generated feedback and design performance: The role of self-efficacy in a quasi-experimental study. *International Journal of Instruction*, *20*(1), 317-332.

INTRODUCTION

The critique, often called the crit, has long served as the primary vehicle for formative feedback in design education. In architecture, graphic design, and advertising programs alike, instructors gather students' work, comment on it publicly or individually, and expect students to revise in light of what was said. This model carries real pedagogical weight: it exposes learners to expert evaluation, socializes them into professional norms, and can spark generative dialogue about choices that might otherwise go unexamined (Schon, 1987). Yet what it rarely manages is to deliver feedback while students can still act on it meaningfully within the same project cycle. By the time a class meets for formal critique, the work is typically far advanced, and studio schedules rarely permit a full restart.

This temporal constraint helps explain growing interest in AI-assisted feedback tools in creative education. Systems capable of providing structured, criterion-referenced commentary on student submissions at any time, and of doing so repeatedly within a single project, offer a different feedback rhythm from the weekly or biweekly crit (Meyer et al., 2024; Schmitt-Fumian et al., 2025). Whether that different rhythm produces measurably better learning outcomes in design education remains an open question. While beneficial effects of AI-generated feedback have been reported in writing and language-learning contexts (Escalante et al., 2023; Hao & Razali, 2026; Lu, 2025), design education involves evaluative dimensions such as aesthetic coherence, conceptual originality, and visual rhetoric that are less tractable to automated analysis than grammatical correctness, and direct experimental comparisons in design programs remain limited (Kasneci et al., 2023; Meron & Tekmen Araci, 2023).

It is also important to distinguish between two educational roles of AI. One is AI as a creative support tool that helps generate ideas, suggest directions, or automate routine design tasks. The other is AI as a structured feedback tool that evaluates work against criteria, identifies areas for improvement, and recommends next steps. These roles involve different pedagogical mechanisms, different risks, and different validation demands (Alamasi & Asfour, 2026; Meron & Tekmen Araci, 2023). The present study focuses specifically on the feedback role.

In advertising design programs, the stakes are especially clear. Students must make decisions about typography, layout, image use, and verbal-visual integration under conditions of genuine uncertainty about what will communicate most effectively. Teachers' critiques provide one form of guidance; peer feedback, self-assessment, and iterative testing provide others. Yet the standard studio model lacks a built-in mechanism for students to receive structured feedback on work in progress, such as early concept sketches or layout drafts, before those decisions are locked into a final submission. AI-assisted feedback tools designed to promote iterative revision rather than simply evaluate finished products may help address that gap.

Three research questions guided this study: (RQ1) Did the AIFF group's design quality differ from the TLC group's at semester end after controlling for pretest performance? (RQ2) Were there corresponding differences in creative performance as assessed by independent judges? (RQ3) Did the two conditions differ in students' design self-

efficacy? A fourth, exploratory question asked whether iteration frequency within the AIFF group predicted individual design quality gains.

Context and Review of Literature

The Pedagogical Architecture of Feedback in Design Education

Feedback research in higher education has converged on several widely accepted principles: effective feedback is timely, specific, and actionable; it supports learners in self-regulating their work rather than merely signaling a grade; and it is most productive when students can respond to it through revision rather than receiving it as a terminal verdict (Black & Wiliam, 1998; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Sadler (1989) argued that learners must recognize the gap between current and desired performance and possess strategies for closing it. Hattie and Timperley (2007) further mapped feedback across task, process, self-regulation, and self levels, emphasizing the importance of the middle two. Hattie (2009) later synthesized a broad evidence base showing that feedback is among the strongest influences on student achievement, although not all feedback is equally effective. Panadero and Lipnevich (2022) and Kluger and DeNisi (1996) both stressed that feedback effectiveness depends on cognitive, motivational, and behavioral conditions operating together.

Wisniewski et al. (2020) updated this evidence base with a meta-analysis showing that feedback effects are strongest when learners can act iteratively on commentary and possess strong task-specific confidence. This link between iterative behavior and self-efficacy is especially relevant to design education, where projects are developed through multiple rounds of choice, revision, and evaluation.

In design education, these principles take a distinctive form. Design tasks are open-ended, not closed-form problems with one correct answer. Schon (1987) described design education as a reflective practicum in which instructors demonstrate ways of framing problems rather than transmitting fixed solutions. Empirical work on studio critique shows that students make better sense of commentary when critique is organized dialogically rather than as one-way judgment (Tahsiri, 2020). Expertise level further complicates matters because novice and expert designers often decompose design problems differently during ideation (Ho, 2001). Relational dynamics matter as well: in architecture and design juries, power asymmetries can suppress student voice and turn critique into compliance rather than exploration (Webster, 2006). More broadly, design critique involves multiple feedback relations among faculty, peers, and the self, and students engage more productively when comments are specific enough to guide subsequent revision (Travis, 2013).

AI-Mediated Feedback in Creative and Related Educational Contexts

Large language models have entered creative education faster than rigorous outcome research has kept pace. In a controlled design-education study, Schmitt-Fumian et al. (2025) found that generative AI feedback supported ideation and iterative refinement during brainstorming, although the authors also noted that AI feedback involved less relational scaffolding than human educator feedback. Meron and Tekmen Araci (2023)

similarly discussed the potential of ChatGPT in design education while emphasizing the generic nature of much AI-produced content and the need for human judgment in curriculum design.

Research in adjacent educational domains provides stronger evidence for feedback outcomes. Meyer et al. (2024) reported that AI-generated feedback increased secondary students' text revision, motivation, and positive emotions, while Escalante et al. (2023) found that AI-generated and human-tutor writing feedback produced comparable learning outcomes with different perceived strengths. In the International Journal of Instruction, Hao and Razali (2026) showed that AI-assisted peer feedback improved self-regulated learning, self-efficacy, and motivation in English academic writing, and Lu (2025) found that combining peer assessment with AI review improved undergraduate English writing and evaluation ability. Together, these studies suggest that AI feedback may be particularly useful when it is repeated, structured, and embedded in active revision.

At the same time, the quality of AI feedback depends heavily on prompt design and system architecture. Kasneci et al. (2023) argued that LLM-based educational tools are most promising in high-frequency, low-stakes formative contexts rather than in holistic high-stakes assessment. Liang et al. (2023) also showed that LLM-generated feedback can be detailed and useful, but remains less context-sensitive than expert human feedback on ambiguous, high-judgment tasks. In the Anatolian Journal of Education, Dertli and Yildiz (2025) demonstrated that prompt engineering substantially shaped the quality of AI-generated educational outputs, while Tangkui (2026) highlighted teacher readiness as an important implementation factor when AI tools are introduced into instructional settings. Although these studies come from different subject areas, they reinforce two points relevant to the present study: prompt quality matters, and implementation conditions matter.

Related work on automated feedback has further suggested that such tools can contribute to students' feedback literacy when learners are supported to interpret and act on the comments they receive (Tubino & Adachi, 2022; Molloy et al., 2020). Students who actively seek commentary rather than passively receive it may be especially likely to benefit, a distinction that is highly relevant in AI-mediated environments where each feedback cycle must often be initiated by the learner (Carless & Young, 2024).

Self-Efficacy and Iterative Revision

Self-efficacy, the belief that one can execute a specific course of action, is among the most reliable predictors of complex performance, including creative performance (Bandura, 1997). In design education, efficacy beliefs are especially relevant because they can diverge from measured ability and may shape behavior more directly than ability alone (Schunk & DiBenedetto, 2020). Creative self-efficacy predicts creative behavior independently of general self-efficacy and cognitive ability (Tierney & Farmer, 2002). Students with stronger design self-efficacy experiment more freely, set harder goals, and persist longer through revision than peers at comparable skill levels (Carmel-Gilfilen & Portillo, 2010).

Bandura (1997) identified mastery experiences as the most potent source of self-efficacy. A feedback condition that allows students to revise, check, and revise again may therefore contribute to stronger self-efficacy by generating more frequent mastery experiences. However, in the present study self-efficacy was examined as an outcome variable, not as a tested mediator or moderator. The design does not permit causal pathway claims about whether self-efficacy mediated performance gains.

Research gap and present study

No published study has directly compared conventional teacher critique with a structured AI-assisted iterative feedback model in an advertising design course using independent outcome measures and a design that controls for pretest differences. Prior research has examined AI feedback in general educational settings (Escalante et al., 2023; Hao & Razali, 2026; Meyer et al., 2024; Lu, 2025) and explored AI applications in creative domains more broadly (Alamasi & Asfour, 2026; Meron & Tekmen Araci, 2023), but the specific comparison between AI-driven iterative feedback and traditional studio critique has not been tested in a design-studio context where both conditions remain pedagogically realistic. Given the accumulated evidence on feedback timing, iterative revision, and the theoretical relevance of self-efficacy, such a comparison is well motivated by theory and practice. In the present study, self-efficacy was examined as an outcome variable rather than as a tested mediator or moderator.

METHOD

Research Design

A quasi-experimental pretest-posttest comparison group design was employed, consistent with the logic of nonequivalent group comparisons described by Shadish et al. (2002). Because random assignment of individual students was not logistically feasible, the unit of assignment was the intact class. Two sections of a third-year Advertising Design Studio course at a regional comprehensive university in southern China were recruited. Instructors, course content, assessment criteria, and the total amount of instructor contact time were held constant across conditions; the primary difference between conditions was the feedback mechanism available to students following each project checkpoint.

Ethical clearance was obtained from the university's Research Ethics Committee prior to data collection. All participants signed informed consent forms, were debriefed at semester end, and were given access to whichever feedback resources they chose after the study concluded.

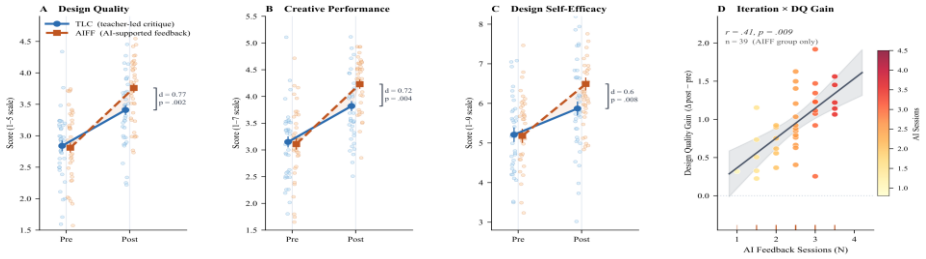

Figure 1
Experimental design overview
Note. TLC = Teacher-Led Critique (n =38); AIFF=AI−supported Iterative Feedback(n= 39). Diamonds (◆) mark the three project checkpoints at which AIFF students could initiate AIDesignCoach consultations between Weeks 4 and 15. Grey boxes at bottom indicate the three outcome constructs with their respective measurement instruments and reliability coefficients.

Participants

Seventy-seven students completed all study components: 38 in the TLC condition and 39 in the AIFF condition. Participants were drawn from two intact classes enrolled in the same third-year Advertising Design Studio course during the Spring 2025 semester. Mean age was 20.8 years (SD = 0.93). Both sections were taught by the same instructor, who had nine years of teaching experience in commercial art and advertising design. Independent-samples t tests confirmed no significant group differences on any of the three baseline measures; all t values were below 0.35 and all p values exceeded 0.73.

Conditions

Teacher-led critique (TLC). Students in this condition participated in instructor-led critique sessions at each major project checkpoint, scheduled approximately every four to five weeks during the active project period. For each session, the instructor reviewed submitted drafts, provided written and verbal commentary, and held 10-15 minute individual conferences per student pair. Critique sessions followed the usual studio format: work was pinned up or projected, strengths and weaknesses were discussed, and students were expected to produce a revised version before the subsequent checkpoint. Across the semester, TLC students received feedback on approximately four formal occasions.

AI-supported iterative feedback (AIFF). Students in this condition had individual access to AIDesignCoach, a custom system built on the Qwen API using a three-layer prompt architecture (Qwen Team, 2025). The system was organized around (a) a conceptual layer covering communicative intent, target audience relevance, and core visual idea; (b) an execution layer covering typography, color, spatial hierarchy, and image-text integration; and (c) a revision layer providing ranked next-step recommendations. Each feedback cycle required students to upload or describe a specific draft stage and indicate which layer of feedback they were seeking.

The system was designed to return a maximum response length of approximately 450 words per session in a structured format. Students could initiate as many feedback

sessions as they wished. The instructor reviewed AIFF transcripts weekly to ensure students were not receiving technically erroneous advice, but did not intervene in the content of the feedback itself. Across the semester, the mean number of AIFF sessions per student was 2.43 (SD = 0.71), with a minimum of 1 and a maximum of 4.

Both groups attended the same weekly studio sessions (3 hours per week) and had identical access to the instructor for procedural questions. The AIFF condition therefore differed from the TLC condition not only in feedback source but also in feedback accessibility and iteration opportunities.

Instruments

Design quality. An eight-criterion analytic rubric was developed for the study in consultation with four advertising design professionals from the university's industry advisory panel. Criteria covered conceptual clarity, target audience alignment, visual hierarchy, typographic execution, color and compositional balance, image-text integration, originality of concept, and overall communicative effectiveness. Each criterion was scored on a five-point scale and the composite score was rescaled to a 1-5 metric. Two raters, both practicing advertising art directors with graduate-level teaching experience, scored each student's submitted work independently and blind to condition assignment. ICC(2,1) for the two-rater composite was 0.89 (95% CI [0.85, 0.92]). In cases where the two raters' composite scores differed by more than 0.50, a third rater adjudicated.

Creative performance. Creative performance was assessed using the Consensual Assessment Technique (Amabile, 1982). Three judges, an award-winning advertising creative director, a senior graphic design educator, and an advertising studies researcher, rated all final project compositions on a seven-point scale after sorting works from least to most creative. Judges worked blind to condition assignment and to whether work was a pretest or posttest submission. ICC(2,1) for the three-rater composition was 0.86 (95% CI [0.81, 0.90]), consistent with prior CAT reliability research (Kaufman et al., 2009; Stefanic & Randles, 2015).

Design self-efficacy. A 12-item design self-efficacy scale was adapted from Bandura's (1997) guidelines to reflect specific competence beliefs relevant to advertising design students, such as confidence in creating an effective visual hierarchy or generating multiple concept directions. Items were rated on a nine-point scale from 1 (cannot do at all) to 9 (highly certain can do). The scale demonstrated acceptable internal consistency at pretest (alpha = 0.81) and posttest (alpha = 0.84).

Iteration frequency. For the AIFF condition only, the number of AIDesignCoach sessions initiated by each student across the semester was recorded from system logs. This variable served as a predictor in an exploratory within-group analysis of design quality gains.

Procedure

Data collection unfolded in four stages. At the start of the semester (Week 1), all participants completed pretest assessments: the design self-efficacy scale and

submission of a pilot mini-project rated blind by rubric and CAT judges. Condition assignment began in Week 2. The TLC class continued with the standard studio format, while the AIFF class received orientation training on AIDesignCoach over two 45-minute sessions in Weeks 2 and 3, covering how to frame queries at each feedback layer and how to document feedback received. During Weeks 4-15, both classes completed three major projects, with TLC students receiving critique at each project's checkpoint and AIFF students having open access to AIDesignCoach. Posttest assessments were administered during Week 16 following the same protocol as the pretest.

Data Analysis

For each of the three primary outcome variables, a univariate analysis of covariance (ANCOVA) was conducted with group (AIFF vs. TLC) as the between-subjects factor and the corresponding pretest score as the covariate. Homogeneity of regression slopes was tested for each model; no significant interactions between condition and pretest score were found for any outcome, confirming that ANCOVA assumptions were satisfied. Effect sizes were reported as partial eta squared from ANCOVA and Cohen's d from observed posttest means and pooled posttest standard deviations. Observed posttest means and raw gain scores are presented for descriptive comparison, whereas adjusted posttest means from ANCOVA provide the inferential basis for between-group comparisons after controlling for baseline differences. For the exploratory analysis, a bivariate Pearson correlation was calculated between iteration frequency and design quality gain scores within the AIFF group. All analyses were conducted in SPSS version 27 with alpha = 0.05.

FINDINGS

Preliminary Analyses

Pretest scores across the two conditions did not differ significantly on any of the three outcomes (all p values > 0.70), confirming pretest equivalence. Levene's test for equality of variances was non-significant for all three posttest outcomes, indicating comparable variance across conditions. ANCOVA assumption testing confirmed that the interaction between pretest scores and group assignment was not significant in any of the three models (all interaction p values > 0.15), supporting the homogeneity-of-regression-slopes requirement.

Descriptive Statistics

Table 1 presents observed means and standard deviations for all three outcomes at pretest and posttest, together with adjusted posttest means estimated from ANCOVA. Both groups improved on all three measures over the course of the semester. In terms of observed raw gains, AIFF students improved more on each measure: design quality increased by 0.95 points in the AIFF group and 0.57 in the TLC group; creative performance increased by 1.12 versus 0.67; and self-efficacy increased by 1.31 versus 0.66. These observed values are descriptive. The adjusted posttest means reported in Table 1 provide the basis for the inferential comparison after controlling for pretest performance.

Table 1
Descriptive statistics for design quality, creative performance, and design self-efficacy by group and time point

Outcome	Group	Pretest M (SD)	Posttest M (SD)	Adjusted Posttest M (SE)
Design Quality (1–5 scale)	TLC (n = 38)	2.84 (0.51)	3.41 (0.48)	3.42 (0.08)
	AIFF (n = 39)	2.81 (0.49)	3.76 (0.44)	3.75 (0.08)
Creative Performance (1–7 scale)	TLC (n = 38)	3.15 (0.64)	3.82 (0.59)	3.84 (0.09)
	AIFF (n = 39)	3.11 (0.66)	4.23 (0.54)	4.21 (0.09)
Design Self-Efficacy (1–9 scale)	TLC (n = 38)	5.21 (1.14)	5.87 (1.08)	5.89 (0.16)
	AIFF (n = 39)	5.18 (1.11)	6.49 (1.02)	6.47 (0.16)

Note. Pretest and posttest M values are observed descriptive means. Adjusted posttest means are estimated marginal means from ANCOVA with the corresponding pretest score entered as a covariate. TLC = teacher-led critique; AIFF = AI-supported iterative feedback.

ANCOVA Results

Table 2 presents the ANCOVA summary for each outcome. For design quality, the pretest covariate accounted for a substantial portion of posttest variance, $F(1, 74) = 23.41$, $p < .001$, partial eta squared = 0.240. Group assignment remained significant after this adjustment, $F(1, 74) = 9.87$, $p = 0.002$, partial eta squared = .118. The AIFF group's adjusted posttest mean was 0.33 points higher than the TLC group's on the 1-5 scale.

Creative performance showed the same pattern. Prior performance again predicted posttest scores, $F(1, 74) = 19.87$, $p < 0.001$, partial eta squared = 0.212, and the group difference held after controlling for it, $F(1, 74) = 8.94$, $p = 0.004$, partial eta squared = 0.108. AIFF students averaged 0.37 points higher than TLC peers on the 1-7 CAT scale.

Self-efficacy showed the largest pretest-posttest covariance of the three outcomes, $F(1, 74) = 31.22$, $p < 0.001$, partial eta squared = 0.297. The group effect nonetheless remained significant, $F(1, 74) = 7.48$, $p = 0.008$, partial eta squared = 0.092. On the nine-point scale, AIFF students sat 0.58 points above TLC peers. Cohen's d of 0.60 for self-efficacy indicates a moderate descriptive difference in posttest confidence.

Table 2
Analysis of covariance summary for design quality, creative performance, and design self-efficacy

Outcome	Source	SS	df	MS	F	p	η^2p
Design Quality	Pretest (covariate)	2.84	1	2.84	23.41	< 0.001	0.240
	Group	1.19	1	1.19	9.87	0.002	0.118
	Error	8.98	74	0.121			
Creative Performance	Pretest (covariate)	7.63	1	7.63	19.87	< 0.001	0.212
	Group	3.43	1	3.43	8.94	0.004	0.108
	Error	28.42	74	0.384			
Design Self-Efficacy	Pretest (covariate)	38.91	1	38.91	31.22	< 0.001	0.297
	Group	9.33	1	9.33	7.48	0.008	0.092
	Error	92.26	74	1.247			

Note. All F values for group effects satisfied the homogeneity-of-regression-slopes assumption (all interaction p values > 0.15).

Table 3 provides effect sizes and 95% confidence intervals for all three outcomes. Cohen's d values ranged from 0.60 to 0.77. These effect sizes should be interpreted with appropriate caution given the quasi-experimental design: they describe observed between-group differences at posttest and do not by themselves establish that AI alone caused the improvements.

Table 3
Effect sizes and 95% confidence intervals for group differences at posttest

Outcome	AIFF M	TLC M	Cohen's d	95% CI
Design Quality	3.76	3.41	0.77	[0.31, 1.22]
Creative Performance	4.23	3.82	0.72	[0.26, 1.17]
Design Self-Efficacy	6.49	5.87	0.60	[0.14, 1.04]

Note. Cohen's d computed using pooled posttest SD. Confidence intervals for d estimated using the noncentral t distribution method.

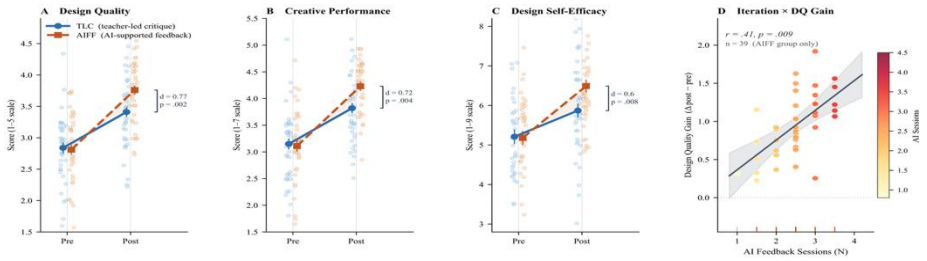

Figure 2
Pre- and posttest mean scores by condition, with iteration frequency effects on design quality (aiff group)

Note. Panels A–C display group means (± 1 SE error bars) at pretest (Week 1) and posttest (Week 16) for design quality (1–5 scale), creative performance (1–7 scale), and design self-efficacy (1–9 scale) respectively. TLC =teacher−ledcritique(filled circles, solid lines, blue); AIFF=AI−supported iterative feedback (filled squares, dashed lines, orange). Cohen's d effect sizes are annotated at each posttest data point; **p < 0.01, two-tailed. Panel D plots individual AIFF participants' total feedback session frequency against design quality gain score posttest−pretest, with ordinary least squares regression line.

Exploratory Analysis: Iteration Frequency

Within the AIFF group, the number of AIDesignCoach sessions per student ranged from 1 to 4 (M = 2.43, SD = 0.71). A Pearson correlation between iteration frequency and the design quality gain score (posttest minus pretest) was significant and positive, r(37) = 0.41, p = 0.009. Students who engaged in more feedback cycles showed larger gains in design quality over the semester. This correlation is consistent with the idea that more feedback cycles may support greater improvement, but its correlational nature precludes causal attribution. It is also possible that students with stronger motivation or greater willingness to engage with AI tools both used the system more often and improved more.

DISCUSSION

Summary of Findings

This quasi-experimental study compared a structured AI-supported iterative feedback condition (AIFF) with conventional teacher-led critique (TLC) on design quality, creative performance, and design self-efficacy among third-year advertising design students. After controlling for pretest performance, the AIFF group showed higher adjusted posttest scores on all three outcomes, with descriptive effect sizes ranging from moderate ($d = 0.60$ for self-efficacy) to large ($d = 0.77$ for design quality). Within the AIFF group, iteration frequency was positively associated with design quality gains, suggesting that the number of feedback cycles may be relevant to the learning process. Taken together, the findings are consistent with the interpretation that a condition combining AI-mediated commentary, greater feedback availability, and more opportunities for iterative revision was associated with stronger outcomes, although the quasi-experimental design does not allow the independent effects of AI, feedback frequency, novelty, or motivation to be separated.

Interpretation in Light of Prior Research

The observed pattern aligns with prior research on feedback timing and revision. Wisniewski et al. (2020) found that feedback effects are strongest when learners can act iteratively on commentary, and the AIFF condition was explicitly organized around that principle. The positive association between iteration frequency and design quality gains within the AIFF group further supports the view that repeated feedback opportunities mattered. At the same time, the present comparison should be interpreted as a contrast between two broader feedback conditions rather than as a clean test of AI alone. Relative to teacher-led critique, the AIFF condition also offered more immediate access to commentary and more opportunities for repeated revision.

The self-efficacy findings are noteworthy because they indicate that students in the AIFF condition reported higher design self-efficacy at posttest than their peers in the TLC condition. One plausible interpretation is that timely, criterion-referenced commentary supported more frequent mastery experiences during revision, which would be consistent with Bandura's (1997) framework and with adjacent AI-feedback findings in writing contexts (Hao & Razali, 2026; Lu, 2025). However, self-efficacy was examined here only as an outcome variable, not as a tested mediator or moderator. The present data therefore do not establish a mediation pathway.

The design quality and creative performance results are also consistent with emerging evidence that AI-mediated feedback can be useful in creative learning environments (Schmitt-Fumian et al., 2025). Yet the present findings are best interpreted cautiously: the observed advantage may reflect a combined contribution of AI mediation, feedback frequency, convenience, and learner engagement rather than a uniquely AI-specific effect.

Theoretical Implications

The study extends feedback theory in creative education in two modest but meaningful ways. First, it supports the argument that the temporal dimension of feedback, including its frequency and proximity to revision opportunities, may be especially consequential in studio learning. Second, it suggests that self-efficacy remains an important companion outcome in design pedagogy, even when it is not modeled as a mediating mechanism. Rather than claiming that AI feedback itself generated self-efficacy, the more defensible conclusion is that a feedback ecology characterized by timely, repeated, criterion-referenced commentary was associated with stronger confidence and stronger performance.

Practical Implications

For design educators, the findings suggest that integrating structured AI feedback systems into studio pedagogy may enhance student learning when those systems are used to support revision rather than replace critique. The AIFF system used in this study was relatively simple, a three-layer prompt architecture built on a general-purpose LLM, yet it was associated with measurable benefits. Educators interested in adopting similar systems should consider providing clear guidance on how students should formulate prompts, monitoring outputs for accuracy and pedagogical suitability, and integrating AI feedback with human critique rather than substituting one for the other.

For educational technology developers, the results reinforce the importance of systems that support multiple feedback cycles throughout the creative process. Findings from AJE also suggest that prompt engineering and implementation readiness can shape educational value in substantial ways (Dertli & Yildiz, 2025; Tangkui, 2026). Systems intended for design education should therefore prioritize transparent criteria, revision-oriented feedback, and user guidance that promotes critical interpretation rather than passive acceptance.

LIMITATIONS

Several limitations should be acknowledged. First, the quasi-experimental design precludes strong causal claims. Although pretest equivalence was established and statistical controls were applied, unmeasured confounding variables may have contributed to the observed differences. In particular, the design does not allow the contribution of AI itself to be disentangled from the effects of greater feedback availability, technological novelty, or possible differences in student motivation and feedback-seeking behavior.

Second, the study was conducted in a single university with a specific cultural and institutional context. The effectiveness of AI-supported feedback systems may vary across educational settings, student populations, and design disciplines.

Third, the AIFF system was built on a single LLM-based infrastructure. Different models and prompt architectures may produce different results, and rapid changes in model capabilities limit generalizability across versions.

Fourth, the study did not examine long-term effects. It remains unclear whether the observed benefits would persist beyond the intervention period or transfer to other design contexts.

Fifth, the study focused on quantitative outcomes and did not explore qualitative aspects of the feedback experience, such as students' trust in AI feedback, their criteria for accepting or rejecting suggestions, or their emotional responses to automated critique. Future studies should combine outcome measures with interviews, trace data, or other process evidence.

Finally, future studies should test whether self-efficacy functions as a mediator or moderator of performance gains rather than treating it only as an outcome.

CONCLUSION

This quasi-experimental study examined whether an AI-supported iterative feedback condition was associated with differences in design quality, creative performance, and self-efficacy relative to conventional teacher-led critique in an advertising design course. Students in the AIFF condition showed higher posttest performance on all three outcomes, with descriptive effect sizes ranging from moderate to large.

These findings suggest that a feedback condition combining AI-mediated commentary, greater feedback availability, and structured opportunities for iterative revision may support learning in creative education contexts. The study therefore contributes to ongoing discussions of how feedback timing and accessibility shape revision processes in studio-based learning. It also indicates that self-efficacy warrants continued attention in this context, although in the present study it was examined only as an outcome and not as a tested mediating mechanism.

At the same time, the findings should be interpreted cautiously. Because the design was quasi-experimental, alternative explanations such as novelty effects, stronger engagement with a new technology, differences in feedback-seeking behavior, or other motivational factors cannot be ruled out. The present study therefore should not be read as showing that AI alone was responsible for the observed differences.

AI-supported feedback systems are best understood here as pedagogical complements rather than replacements for the relational and contextual expertise of human educators. Their potential value lies in extending the reach, timing, and frequency of structured feedback while preserving the dialogic strengths of studio teaching.

FUNDINGS

This research was funded by Chongqing Municipal Education Commission, grant number KJQN202503101, KJQN202403128 and "Mayue" Engineering Project for Student Technological Innovation of Chongqing Polytechnic University of Electronic Technology, Grant No. 25XJXSCX20.

REFERENCES

Alamasi, R., & Asfour, O. S. (2026). Applications of Generative AI in Architectural Design Education: A Systematic Review and Future Insights. *Digital, 6*(1), 6. https://doi.org/10.3390/digital6010006

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology, 43(*5), 997. https://doi.org/10.1037/0022-3514.43.5.997

Bandura, A. (1997). Self-efficacy: The exercise of control (Vol. 11). Freeman.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7 – 74. https://doi.org/10.1080/0969595980050102

Carless, D. (2023). Teacher feedback literacy, feedback regimes and iterative change: towards enhanced value in feedback processes. Higher Education Research & Development, 42(8), 1890-1904. https://doi.org/10.1080/07294360.2023.2203472

Carless, D., & Young, S. (2024). Feedback seeking and student reflective feedback literacy: A sociocultural discourse analysis. *Higher Education, 88*(3), 857-873. https://doi.org/10.1007/s10734-023-01146-1

Carmel-Gilfilen, C., & Portillo, M. (2010). Developmental trajectories in design thinking: an examination of criteria. *Design studies, 31*(1), 74-91. https://doi.org/10.1016/j.destud.2009.06.004

Dertli, Z. G., & Yildiz, B. (2025). The Use of Prompt Engineering in Creating Mathematical Modelling Activities with Artificial Intelligence Tool ChatGPT. *Anatolian Journal of Education, 10*(1), 59-80. https://doi.org/10.29333/aje.2025.1015a

Escalante, J., Pack, A., & Barrett, A. (2023). Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Hongxia, H., & Razali, A. (2026). Impact of AI-Assisted peer feedback on efl students' self-regulated learning, self-efficacy, and motivation in English academic writing. *International Journal of Instruction, 19*(2), 163-182. https://doi.org/10.29333/iji.2026.1929a

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement. routledge.*

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Ho, C. H. (2001). Some phenomena of problem decomposition strategy for design thinking: differences between novices and experts. *Design Studies, 22*(1), 27-45. https://doi.org/10.1016/S0142-694X(99)00030-7

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences, 103,* 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kaufman, J. C., Baer, J., & Cole, J. C. (2009). Expertise, domains, and the consensual assessment technique. *The Journal of creative behavior, 43*(4), 223-233. https://doi.org/10.1002/j.2162-6057.2009.tb01316.x

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin, 119*(2), 254. https://doi.org/10.1037/0033-2909.119.2.254

Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., ... & Zou, J. (2024). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI, 1*(8), AIoa2400196. https://doi.org/10.1056/AIoa2400196

Lu, H. (2025). Effect of peer assessment and artificial intelligence review on English writing and evaluation ability. *International Journal of Instruction, 18*(4), 517-530. https://doi.org/10.29333/iji.2025.18428a

Meron, Y., & Araci, Y. T. (2023). Artificial intelligence in design education: evaluating ChatGPT as a virtual colleague for post-graduate course development. *Design Science, 9*, e30. https://doi.org/10.1017/dsj.2023.28

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence, 6,* 100199. https://doi.org/10.1016/j.caeai.2023.100199

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education, 45*(4), 527-540. https://doi.org/10.1080/02602938.2019.1667955

Nicol, D. J., & Macfarlane‐Dick, D. (2006). Formative assessment and self‐regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35,* 100416. https://doi.org/10.1016/j.edurev.2021.100416

Qwen Team. (2025). *Qwen documentation.* https://qwen.readthedocs.io

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science, 18*(2), 119-144. https://doi.org/10.1007/BF00117714

Schmitt-Fumian, T. K., Tauscher, S., & Thoring, K. (2025). AI vs. human: exploring the potential of generative AI as a feedback tool to support ideation in design education. *Proceedings of the Design Society, 5,* 429-438. https://doi.org/10.1017/pds.2025.10057

Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary educational psychology, 60,* 101832. https://doi.org/10.1016/j.cedpsych.2019.101832

Reichardt, C. S. (2002). *Experimental and quasi- experimental designs for generalized causal inference.* https://doi.org/10.1086/345281

Stefanic, N., & Randles, C. (2015). Examining the reliability of scores from the consensual assessment technique in the measurement of individual and small group creativity. *Music Education Research, 17*(3), 278-295. https://doi.org/10.1080/14613808.2014.909398

Tahsiri, M. (2020). Dialogue in the studio: Supporting comprehension in studio-based architectural design tutorials. *Art, Design & Communication in Higher Education, 19*(2), 149-165. https://doi.org/10.1386/adch_00020_1

Tangkui, R. (2026). AI Integration in Primary Science Instruction: A Technology Acceptance Model (TAM)-Based Study on Teachers' Readiness to Use ChatGPT. *Anatolian Journal of Education, 11*(1), 39-52. https://doi.org/10.29333/aje.2026.1113a

Tierney, P., & Farmer, S. M. (2002). Creative self-efficacy: Its potential antecedents and relationship to creative performance. *Academy of Management journal, 45*(6), 1137-1148. https://doi.org/10.2307/3069429

Travis, S. (2013). The design critique: Faculty to student, student to student, student to self. *The International Journal of Design Education, 6*(3), 77. https://doi.org/10.18848/2325-128X/CGP/v06i03/38422

Tubino, L., & Adachi, C. (2022). Developing feedback literacy capabilities through an AI automated feedback tool. *ASCILITE Publications,* e22039-e22039. https://doi.org/10.14742/apubs.2022.39

Webster, H. (2006). Power, freedom and resistance: Excavating the design jury. *International Journal of Art & Design Education, 25*(3), 286-296. https://doi.org/10.1111/j.1476-8070.2006.00495.x

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in psychology, 10,* 487662. https://doi.org/10.3389/fpsyg.2019.03087

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology, 25*(1), 82-91. https://doi.org/10.1006/ceps.1999.1016